

Levels of Evidence for the Recommendations in Institute of Education Sciences Practice Guides

Level of Evidence	Description
Strong	<p>In general, characterization of the evidence for a recommendation as <i>strong</i> requires both studies with high internal validity, i.e., studies whose designs can support causal conclusions, as well as studies with high external validity, i.e., studies that in total include enough of the range of participants and settings on which the recommendation is focused to support the conclusion that the results can be generalized to those participants and settings. Strong evidence for this practice guide is operationalized as:</p> <ul style="list-style-type: none"> • A systematic review of research that generally meets the standards of the <i>What Works Clearinghouse</i> and supports the effectiveness of a program, practice, or approach and no contradictory evidence of similar quality; OR • Several well-designed, randomized, controlled trials or well-designed quasi-experiments that generally meet the standards of the <i>What Works Clearinghouse</i> and support the effectiveness of a program, practice, or approach and no contradictory evidence of similar quality; OR • One large well-designed, randomized, controlled, multi-site trial that meets the standards of the <i>What Works Clearinghouse</i> and supports the effectiveness of a program, practice, or approach and no contradictory evidence of similar quality; • For assessments, evidence of reliability and validity that meets The Standards for Educational and Psychological Testing.³
Moderate	<p>In general, characterization of the evidence for a recommendation as <i>moderate</i> requires studies with high internal validity but moderate external validity or studies with high external validity but moderate internal validity. In other words, moderate evidence is derived from studies that support strong causal conclusions but where generalization is uncertain, or studies that support the generality of a relationship but where the causality is uncertain. Moderate evidence for this practice guide is operationalized as:</p> <ul style="list-style-type: none"> • Experiments or quasi-experiments generally meeting the standards of the <i>What Works Clearinghouse</i> and supporting the effectiveness of a program, practice, or approach with small sample sizes and/or other conditions of implementation or analysis that limit generalizability and no contrary evidence; OR • Comparison group studies that do not demonstrate equivalence of groups at pretest and therefore do not meet the standards of the <i>What Works Clearinghouse</i> but that (a) consistently show enhanced outcomes for participants experiencing a particular program, practice, or approach and (b) have no major flaws related to internal validity other than lack of demonstrated equivalence at pretest (e.g., only one teacher or one class per condition, unequal amounts of instructional time, highly biased outcome measures); OR • Correlational research with strong statistical controls for selection bias and for discerning influence of endogenous factors and no contrary evidence: • <i>For assessments</i>, evidence of reliability that meets the standards The Standards for Educational and Psychological Testing (1999) but evidence of validity from samples that are not adequately representative of the population on which the recommendation is focused.
Low	<p>In general, characterization of the evidence for a recommendation as <i>low</i> means that the recommendation is based on expert opinion derived from strong findings or theories in related areas and/or expert opinion buttressed by direct evidence that does not rise to the moderate or strong levels. Low evidence is operationalized as evidence not meeting the standards for the moderate or high levels.</p>

³ American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999)